Studies in the Linguistic Sciences
Volume 31, Number 2 (Fall 2001)

CROSS-LANGUAGE SPEECH PERCEPTION IN ADULTS: DISCRIMINATION OF KOREAN VOICELESS STOPS BY ENGLISH SPEAKERS

Sarah J. Shin
University of Maryland Baltimore County
shin@umbc.edu

This study examines the ways in which three classes of alveolar stops in Korean (voiceless 'tense' unaspirated /t*/, voiceless 'lax' slightly aspirated /t/, and voiceless heavily aspirated /t^h/) present different degrees of perceptual difficulty to adult English and Korean listeners. Results show that the /t/-/t^h/ contrast presented the greatest difficulty in perceptual discriminability for the American listeners (61% error rate) while the /t*/-/t/ and the /t*/-/t^h/ contrasts presented relatively easy discriminability with 9% and 3% error rates respectively. Pairwise *t*-test results show that English listeners discriminated the /t*/-/t^h/ contrast significantly better than the /t*/-/t/ contrast, suggesting that a larger difference in VOT between stimulus items increases discriminability. These and other results suggest that English listeners' perception of Korean voicing contrasts is largely determined by phonemic status and to a lesser extent by the magnitude of acoustic difference.

1. Introduction

Cross-language perceptual studies have repeatedly shown that adults typically discriminate all native language contrasts but have difficulty discriminating phonetic contrasts that are not used contrastively in their native language (e.g., Abramson & Lisker 1970; Goto 1971; MacKain, Best, & Strange 1981; Miyawaki et al. 1975; Strange & Jenkins 1978; Tees & Werker 1984; Werker 1991; Werker & Logan 1985; Werker & Tees 1984). The difficulty in discriminating nonnative contrasts has been attributed to the nonphonemic status of these contrasts in the listeners' native language. That is, listeners are sensitive to phonetic differences between sounds when these differences signal a meaningful distinction in their native language, but are apparently insensitive to other phonetic differences. For example, it was found that adult speakers of Japanese have considerable difficulty discriminating between English [r] and [l], a contrast not found in Japanese, whereas English listeners showed highly accurate discrimination of across-category comparisons (Miyawaki et al. 1975).

The effect of phonemic status of the listeners' native language has been reported for a wide range of phonetic properties. For consonants, this includes place, manner, and voicing distinctions. Of particular interest here is cross-language perception study of consonant voicing. The classic acoustic measure of voicing is VOT (Voice Onset Time), which is defined as the difference in time between the release of the stricture (usually, stop closure) and the onset of voicing for the following vowel (Lisker & Abramson 1964). A negative VOT value indicates voicing lead while a positive VOT value indicates voicing lag. Lisker and Abramson (1970) generated 37 synthetic CV stimuli in which the C portion incrementally varied in VOT from -150 to +150 ms. These stimuli were presented to speakers of Thai and English using identification and discrimination paradigms. Thai listeners separated the synthetic labial VOT series into three distinct categories on an identification task (with category boundaries at about –20 msec VOT and +40 msec VOT) and produced two categorical peaks in discrimination corresponding to these two identification boundaries, consistent with the three distinct voicing categories in Thai stops. In contrast, English has a two-way voicing distinction in stops, whose average labial VOT values are 0 msec and 58 msec (Lisker & Abramson 1964). Correspondingly, American subjects produced a single categorical peak at +25 msec VOT and showed no indication of better discrimination for pairs that crossed the Thai -20 msec VOT boundary than for within-category comparisons (Abramson & Lisker 1970).

Although it has been fairly well established that phonemic experience in the listener's native language influences the perception of nonnative contrasts, it has also been noted that some nonnative contrasts present greater difficulty than others (Best 1991; Best, McRoberts, & Sithole 1988; Tees & Werker 1984; Werker & Logan 1985; Werker & Tees 1984). This nonuniform difficulty indicates that phonemic experience alone cannot adequately account for the effects of specific language experience. Aside from the listeners' phonemic experience, it has been claimed that phonetic familiarity and acoustic salience may also underlie differences in perceptual difficulty among nonnative speech contrasts (Polka 1991, 1992; Werker & Logan 1985).

The proposed role of phonetic familiarity is that allophonic experience with the relevant phones may improve discriminability of these phones. For example, in a study of American English speakers' perception of Hindi phoneme contrasts, Tees & Werker (1984) found that English subjects who had had more than five years of Hindi instruction were able to discriminate a Hindi voicing contrast and a Hindi place of articulation (dental-retroflex) contrast, while subjects who had less than one year of instruction could discriminate the voicing but not the place distinction. Tees and Werker suggested that, for the second group of English listeners, allophonic experience with a range of voiced and voiceless variants in English may have been sufficient to enhance discrimination of the Hindi voicing contrast whereas such phonetic familiarity was lacking for the place contrast.

Other studies have shown that efforts to train nonnative speakers to distinguish voicing contrasts have been more successful than training involving place distinctions (Pisoni, Aslin, Perey, & Hennessy 1982), suggesting that voicing distinctions might be inherently more salient. Polka (1991) suggests that the spectral differences found in different places of articulation may be less 'robust' than the temporal differences found in voicing contrasts. Such differences in perceptual difficulty are distinct from those caused by the phonemic status of the listener's native language or the listener's familiarity with nonnative phones and can only be explained by inherent salience of acoustic parameters.

1.1 Best's Perceptual Assimilation Model

Best and her colleagues (Best 1991; Best et al. 1988) have presented a perceptual model that provides an integrated view of how phonemic, phonetic, and acoustic factors may underlie variation in the perception of nonnative speech contrasts. Best et al. (1988) maintain that the listeners' attention is normally focused at the phonemic level during speech perception and that listeners perceptually assimilate nonnative phones to their native phoneme categories whenever possible. Best (1991) proposes four ways in which nonnative phones may be assimilated, and makes several predictions regarding the relative perceptual difficulty of the four assimilation patterns.

In the first pattern, two target phones are assimilated as members of two different native phoneme categories ('two-category' assimilation). This pattern is exemplified by French listeners' perception of the English voicing contrast in voiceless and voiced dental fricatives. French listeners tend to perceive these fricatives as /t/ and /d/ respectively. The second pattern is 'single-category' assimilation, where the two target phones are assimilated as variants of a single native phoneme. An example is English listeners' perception of the Hindi voiceless retroflex-dental contrast, where both consonants are perceived as instances of the same English alveolar stop category (Werker 1991). Best's model predicts that listeners will have the least difficulty discriminating two-category contrasts while having the greatest difficulty discriminating single-category contrasts.

Two other assimilation patterns give rise to intermediate levels of difficulty. In 'category-goodness' assimilation, the nonnative phones may both be assimilated to a single native category, but the listener perceives one phone to be phonetically more similar to the native category than the other nonnative phone. For example, both the Zulu voiceless aspirated velar stop /k^h/ and velar ejective /k'/ are likely to assimilate to English [k^h], but the former is perceived as essentially identical to English [k^h] while the latter is heard as quite discrepant from it. In the fourth pattern, nonnative sounds are highly discrepant from the properties of any native categories and are perceived as nonspeech sounds (i.e., 'nonassimilable' category). The clicks of southern Bantu and Khoisan languages are unlikely to assimilate well to any English phoneme categories because they are unlike any English phonemes. Since the 'nonassimilable' category tokens are perceived as nonspeech sounds, the listener differentiates them by attending to psychoacoustic differences rather than phonemic or phonetic differences.

1.2 The current study

Based on this model then, the current research examines adult English and Korean listeners' perception of the distinctions in phonological types found in Korean alveolar stops in naturally spoken tokens.[1] More specifically, this study attempts to investigate phonemic status, phonetic familiarity, and acoustic salience underlying English listeners' perception of voicing distinctions in Korean, which has a three-way contrast in voiceless stops in initial position. The three Korean stops contrast phonemically and are exemplified by the following minimal triplet:

Type 1: [t*am] 'sweat' (unaspirated, tense or fortis)
Type 2: [tam] 'fence' (slightly aspirated, lax or lenis)
Type 3: [t^ham] 'greed' (heavily aspirated)

Tables 1 through 3 present a summary of previously reported mean VOT values for each of the three categories of Korean stops in word-initial position at labial, alveolar, and velar places of articulation.

Table 1. Mean VOT values (msec) for unaspirated stops in initial position

	p*	t*	k*	mean	# of speakers
Kim (1965)	9	15	13	12	1
Lisker & Abramson (1964)	7	11	20	13	2
Han & Weitzman (1970)	5	10	24	13	2

Table 2. Mean VOT values (msec) for slightly aspirated stops in initial position[2]

	p	t	k	mean	# of speakers
Kim (1965)	23	38	45	35	1
Lisker & Abramson (1964)	20	28	48	32	2
Han & Weitzman (1970)	23	28	52	35	2
Silva (1992)	60	51	71	61	7

Table 3. Mean VOT values (msec) for heavily aspirated stops in initial position

	p^h	t^h	k^h	mean	# of speakers
Kim (1965)	98	92	90	93	1
Lisker & Abramson (1964)	90	96	126	105	2
Han & Weitzman (1970)	117	120	142	126	2

As can be seen in Tables 1 through 3, all three categories of Korean stops lie on the positive side of the VOT continuum, unlike some other three-category languages such as Thai or East Armenian which have a voiced stop (with negative VOT), an unaspirated stop (VOT around 0 ms), and a voiceless stop (with a positive VOT). This presents an interesting perception problem for speakers of English, which has a two-way distinction with a perceptual categorical peak around +25 msec. If VOT is indeed a salient acoustic feature in cross-language perception of stops, as has been claimed by previous research mentioned so far, we should be able to make the following predictions regarding English listeners' perception of Korean stops.

1.3 Perception predictions

In terms of Best's Perceptual Assimilation Model, Korean /t*/ and /t/ (which cross the English phonemic boundary) should be assimilated as two categories,

/d/ and /t/ in English; the same would be true of /t*/ and /t^h/. However, Korean /t/ and /t^h/ would be assimilated as a single category since both of their VOT values fall within the English /t/ category. If phonemic status is the sole factor in explaining perception of the Korean contrasts, then /t*/-/t/ and /t*/-/t^h/ contrasts would be predicted to be equally easy. However, if acoustic salience exerts an influence on perception of Korean phones, then perception of /t*/-/t^h/ and /t*/-/t/ contrasts would be predicted to be different. Based on the claim made by Carney, Widin, & Viemeister (1977) that a larger VOT difference is perceptually more salient than a smaller VOT difference, English listeners would be predicted to better discriminate the /t*/-/t^h/-contrast than the /t*/-/t/-contrast since the former pair is characterized by a larger VOT difference.

2. Method

2.1 Stimuli

A 30-year-old male native speaker of the Seoul dialect of Korean was recorded producing multiple repetitions of CV syllables where C was one of the three Korean alveolar stops and V was /a/. The speaker had been living in the United States for six years. Multiple repetitions of 20 randomized Korean CV syllables including the 3 alveolar stop types and 17 other Korean consonants were recorded by the speaker in a sound-attenuated chamber. The 8 tokens of each stop category were low-pass filtered at 5 KHz and digitized at 10 KHz with 16 bit resolution. Stimulus materials were prepared using a multiple natural exemplar approach (Polka 1991) in which tokens of each stop category are selected so as to minimize across-token differences within that category.

On the basis of acoustic analysis and perceptual judgments by three other native Korean speakers, four tokens of each stop category were selected so that acoustic variations among the four tokens were minimal. Two criteria were used to determine the amount of acoustic variation between tokens: VOT and F0 at the onset of the following vowel. This is done because, in addition to VOT differences between the Korean stop categories, it has been found that F0 following /t*/ is higher (by about 30 Hz) than that following /t/ and F0 following /t^h/ is higher (by about 10 Hz) than that following /t*/ (Han & Weitzman 1970; Silva 1992).

VOT values (measured from the release of the stop closure to the onset of voicing) and F0 at the onset of the following vowel were obtained by using spectrographic and waveform displays. The VOT values and the F0 measures for voiceless unaspirated /t*/, voiceless slightly aspirated /t/, and voiceless heavily aspirated /t^h/ are listed in Tables 4 through 6.

VOT values for the four tokens of each category differed by up to 8 msec for the /t/ and /t^h/ categories and by about 4 msec for /t*/. Test tapes consisted of pairings of the 12 stimuli arranged in a categorial AX task that requires the listener to attend to between-category differences and to ignore within-category variation. (Thus, the categorial AX task used by Polka (1991), among others, is conceptually similar to the name-identity AXB discrimination task used by Best et al. (1988)).

Table 4. VOT and F0 measures for 4 tokens (t*1-t*4) of /t*/

/t*/	VOT (ms)	F0 of V (Hz)
t*1	8.01	128
t*2	9.76	127
t*3	6.93	129
t*4	10.18	129
mean	8.72	128.25

Table 5. VOT and F0 measures for 4 tokens (t1-t4) of /t/

/t/	VOT (ms)	F0 of V (Hz)
t1	37.61	117
t2	29.83	112
t3	36.27	116
t4	31.92	117
mean	33.90	115.5

Table 6. VOT and F0 measures for 4 tokens (t^h 1- t^h 4) of /t^h/

/t^h/	VOT (ms)	F0 of V (Hz)
t^h1	90.51	137
t^h2	93.62	136
t^h3	86.31	134
t^h4	94.40	135
mean	91.21	135.5

The test sequence consisted of 48 12-trial blocks (48 blocks X 12 trials/ block = 576 trials total). Every possible combination of 'same' and 'different' pairs were created except that no token was paired with itself (e.g., t*1 was never paired with t*1). For example, as shown in Fig. 1, the /t*/-category 'same' pairs had 6 combinations which could occur in either order (e.g., t*1-t*2 and t*2-t*1). Corresponding 'same' pairings were created for the /t/ and /t^h/ categories as well. Figure 2 shows an example of 'different' pair combinations between /t/ and /t*/. Sixteen pairs result from such pairings and these are in turn reversed as above. Corresponding 'different' pairings for /t/-/t^h/ and /t*/-/t^h/ were also created. There were 16 repetitions of each of the 18 'same' pair types and 6 repetitions of each of the 48 'different' pair types so as to make the same total number of 'same' and 'different' pairs. Each 12-trial block contained 6 'same' pairs and 6 'different' pairs in a randomized fashion and the entire test sequence consisted of 288 'same' pairs and 288 'different' pairs.

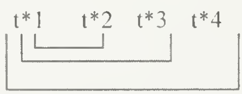

Figure 1. Sample 'same' combinations

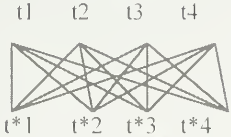

Figure 2. Sample 'different' combinations

The interval between two stimuli in a trial (e.g., between t*1 and t*2 in a 'same' pair trial, or between t1 and t*1 in a 'different' pair trial), or interstimulus interval (ISI), was 1.5 seconds and the interval between trials, or intertrial interval (ITI), was 3 seconds. The interval between blocks was 8 s. A relatively long ISI of 1.5 seconds was chosen because listeners tend to demonstrate phonemic-level processing with a longer ISI and auditory processing with a shorter ISI (Werker & Tees 1984). The Korean syllables on the test tapes were identified by the author to determine whether the stimulus items had been distorted or mislabeled during stimulus processing.

2.2 Subjects

20 American-English speaking undergraduate students at the University of Michigan and 10 native Korean speakers served as subjects. The 8 male and 12 female American subjects ranged from 18 to 36 years of age, with a mean age of 20.6 years. The extent of the American subjects' knowledge of foreign languages ranged from no training at all to several years of classroom instruction in one or more of Spanish, French, Italian, German, Hebrew, and Vietnamese. The control group of 6 male and 4 female Korean subjects were either graduate students at the University of Michigan or their spouses and ranged from 23 to 33 years of age, with a mean age of 27.7 years. Length of stay in the U.S. for the control group ranged from 1.5 years to 6 years, with a mean length of stay of 3.8 years. Subjects were paid for their participation. All subjects reported having normal hearing and having had no phonetics training.

2.3 Procedure

2.3.1 American-English Listeners

American-English subjects were told that they would hear syllables beginning with three different consonants that do not differentiate words in English but are distinctive in other languages. To familiarize subjects with the three-way distinction, two or three examples of each of the three stop categories were played one or more times so that the total number of stimuli played per category was five. The stimuli in this familiarization sequence were not those used in the test sequences. Following the familiarization sequence, a brief 12-trial practice test was given, again using tokens that were not used in the actual test. While the first 6 trials of the practice test had correct answers written on the response sheet, the last 6 trials contained blanks for listeners to write their own responses. Answers for the last 6 trials were given after the practice test was completed and listeners graded their own responses.

The recorded test sequences were presented through Sennheiser HD 222 headphones connected to a Panasonic Audiotape Deck SV-3500 in a sound-attenuated chamber to one to four listeners at a time. Subjects were asked to respond 'same' when they thought the pair contained two instances of the same consonant, and 'different' when they thought the pair contained two different consonants, by writing 'S' in the former case and 'D' in the latter. It was explained that 'same' does not mean 'identical' but rather refers to two instances of the same stop category. During the actual test, two five-minute breaks were given: one after the first 18 blocks, another after the next 18 blocks. During breaks, all listeners were asked to leave the sound-attenuated chamber and move around in the hallway. The entire procedure was approximately one hour and 30 minutes in duration.

2.3.2 Korean Listeners

Instructions to Korean listeners were given by the author in Korean. Korean listeners were told that they would hear syllables beginning with three different consonants in Korean. The same familiarization sequence and the 12-trial practice test were given to the Korean listeners. Subjects were asked to respond 'same' when they thought the pair contained two instances of the same consonant, and 'different' when they thought the pair contained two different consonants, by writing an 'O' (a circle) in the former case and 'X' (an ex) in the latter. Again, explanation was provided that 'same' does not mean 'identical' but rather refers to two instances of the same stop category. Answers for the last 6 trials were given after the practice test was completed and listeners graded their own responses.

3. Results

The mean percent errors of the 10 Korean listeners in the categorial AX task for each of the Korean contrasts are shown in Table 7. The Korean listeners had a 99% accuracy rate, except for one subject who had an accuracy rate of 92%. It should be noted that 4% out of the 5% error rate for the /t*/-/t*/ pair and 0.8% out of the 1.4% error rate for the /t^h/-/t^h/ pair was due to that listener. The source of this particular listener's difficulties is unclear since the length of her stay in the U.S. is 3.5 years, close to the mean length of stay of 3.8 years for the ten Korean subjects. Overall, however, the near-perfect accuracy rate of Korean listeners' perceptual data suggests that the tokens used in the perception test are clear instances of Korean stops.

Table 7. Pooled perceptual results for Korean listeners

Contrast	t*- t*	t- t	$t^h - t^h$	t*-t	t*- t^h	t- t^h
% Error	5.0	0.6	1.4	0.1	0.1	0.5

The mean percent errors of the 20 American English listeners in the categorial AX task for each of the Korean contrasts are shown in Table 8. While discrimination of the /t*/-/t/ contrast and the /t*/-/t^h/ contrast appears to be relatively easy for English-speaking listeners, the /t/-/t^h/ contrast seems to be much more difficult with a greater than chance error rate (with chance being 50%).

Table 8. Pooled perceptual results for American-English listeners

Contrast	t* - t*	t- t	t^h - t^h	t*-t	t*- t^h	t- t^h
% Error	23.7	11.3	22.6	8.5	2.9	60.9

Pairwise *t*-tests were performed, comparing the % errors for the three sets of 'different' pairs. The results showed that the % error for the /t/-/t^h/ pairs was significantly greater than that of both the /t*/-/t/ pairs (t = 17.3, p<0.0001) and the /t*/-/t^h/ pairs (t = 19.001, p<0.001). In addition, the % error for the /t*/-/t/ pair was also significantly greater than that of the /t*/-/t^h/pair (t = 4.9376, p<0.0001). As for the three 'same' categories, it is not clear what caused the large difference in the error rates of the sameness judgments. (But see section 4 for a discussion of some possible factors.)

4. Discussion

In the Introduction, different predictions were outlined regarding the relative differentiability of the three Korean contrasts by English listeners. First, phonemic status as a predictor of perception of the three contrasts was evaluated. Based on Best's Perceptual Assimilation Model, it was predicted that English listeners would clearly distinguish the Korean voiceless unaspirated stops, /t*/, from both slightly aspirated stops, /t/, and heavily aspirated stops, /t^h/. This pattern is well reflected in the perceptual data. While both the /t*/-/t/ and the /t*/-/t^h/ contrasts produced low error rates, the /t/-/t^h/ contrast resulted in a significantly higher error rate.

Recall that if we were to take only phonemic status into account, English listeners' perceptions of /t*/-/t/ and /t*/-/t^h/ contrasts were predicted to be equally difficult. The acoustic salience factor, on the other hand, predicted that English listeners would better discriminate the /t*/-/t^h/ contrast because of a larger VOT difference between /t*/ and /t^h/ than between /t*/ and /t/. Although the difference in error rates between the two pair types is indeed small (less than 6%), the difference is significant in the predicted direction, suggesting that the /t*/-/t^h/ contrast was in fact easier to discriminate. This result supports the idea that the magnitude of difference in VOT, along with the phonemic status of English, influences English listeners' perception of Korean voicing contrasts.

Turning to listeners' performance on the 'same' pair types, the error rates on these pairs ranged from 11% to nearly 24%. This points to a fairly high level of uncertainty in making 'same' judgments which presumably is due in part to the within-category acoustic variation in the four tokens of each category type. This variation arises from the use of natural speech tokens rather than synthetic tokens that provide complete control over acoustic parameters (see also Polka 1991). Since the within-category VOT and F0 differences (Tables 4-6) do not clearly correlate with the level of perceptual difficulty (Table 8), it may be that English listeners (and perhaps the one Korean listener who had an unexpectedly high error rate) were attuned to acoustic differences other than VOT and F0. A future study may determine the relationships between perceptual difficulty and other specific acoustic parameters not examined here, such as formant transitions, total duration of the syllable, vowel duration, vowel mean amplitude, and vowel peak amplitude.

In summary, this study investigated how English-language listeners perceive phonation distinctions among Korean alveolar stops. Predictions based on the phonemic status of these stop categories in English and on the magnitude of the acoustic differences (in selected dimensions) were generally supported by English-listeners' response patterns to the across-category ('different') pair types. However, the factors underlying their responses to within-category ('same') pair-types were not conclusively determined and would require further investigation. Overall, the perceptual data suggest that neither phonemic status nor acoustic salience can individually account for the variability in perception of the Korean contrasts. Rather, a more integrated approach that incorporates these and other acoustic factors seems to be necessary in producing a complete explanation of the manner in which English listeners perceive Korean stop contrasts.

NOTES

[1] In this study, only the alveolar series was investigated due to the large number of trials required in the perception experiment (see also section 2.1).

[2] Silva (1992) did not provide VOT measurements for unaspirated and heavily aspirated stops. For slightly aspirated stops, there is a striking discrepancy between Silva's study and the previous ones, which Silva attributed to differences in measurement procedures. Silva reported that Kim (1965) and Lisker & Abramson (1964) measured VOT from the beginning of the release to the onset of voicing in the following vowel; Han & Weitzman (1970) measured from the beginning of the release to the onset of F1; and Silva (1992) measured from the end of the release to the onset of F2. Since it is often the case that onset of voicing occurs before the onset of F1, which in turn occurs before the onset of F2, Silva reasoned that using each of these three points to mark the end of aspiration would yield values that are progressively longer in each case.

REFERENCES

ABRAMSON, Arthur S., & Leigh LISKER. 1970. Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Science* 569-73. Prague: Academia.

BEST, Catherine T. 1991. The emergence of native-language phonological influences in infants: a perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research* SR-107/108, 1-30.

BEST, Catherine T., Gerald W. MCROBERTS, & Nomathemba M. SITHOLE. 1988. Examination of perception reorganization for non-native speech contrasts:

Zulu click perception by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14.345-60.

CARNEY, Arlene Earley, Gregory P. WIDIN, & Neal F. VIEMEISTER. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62:4.961-71.

GOTO, H. 1971. Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia* 9.317-23.

HAN, M. S., & R. S. WEITZMAN. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica* 22.112-28.

KIM, Chin-W. 1965. On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word* 21.339-59.

LISKER, Leigh, & Arthur S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20.384-422.

——. 1970. The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Science* 563-67. Prague: Academia.

MACKAIN, Kristine S., Catherine T. BEST, & Winifred STRANGE. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2.369-90.

MATTINGLY, I. G., & M. STUDDERT-KENNEDY (eds.). 1991. *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Erlbaum.

MIYAWAKI, Kuniko, Winifred STRANGE, Robert VERBRUGGE, Alvin M. LIBERMAN, James J. JENKINS, & Osamu FUJIMURA. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18.331-40.

PISONI, David B., Richard N. ASLIN, Alan J. PEREY, & Beth L. HENNESSY. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8.297-314.

POLKA, Linda. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89:6.2961-77.

——. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52:1.37-52.

SILVA, David James. 1992. The phonetics and phonology of stop lenition in Korean. Cornell University, Ph.D. dissertation in Linguistics.

STRANGE, Winifred, & James J. JENKINS. 1978. The role of linguistic experience in the perception of speech. In: Walk & Pick (eds.), 125-69.

TEES, Richard C., & Janet F. WERKER. 1984. Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38.579-90.

WALK, R. D., & H. J. PICK, Jr. (eds.). 1978. *Perception and Experience*. New York: Plenum.

WERKER, Janet F. 1991. The ontogeny of speech perception. In: Mattingly & Studdert-Kennedy (eds.), 91-109.

WERKER, Janet F., & John S. LOGAN. 1985. Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37.35-44.

WERKER, Janet F., & Richard C. TEES. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75.1866-78.